

## RESEARCH PAPER

WILEY

Global Ecology  
and BiogeographyA Journal of  
Macroecology

# A simple spatially explicit neutral model explains the range size distribution of reef fishes

Adriana Alzate<sup>1,2,3</sup>  | Thijs Janzen<sup>4,5</sup>  | Dries Bonte<sup>3</sup>  | James Rosindell<sup>6\*</sup> | Rampal S. Etienne<sup>1\*</sup> 

<sup>1</sup>Groningen Institute for Evolutionary Life Sciences, University of Groningen, Groningen, The Netherlands

<sup>2</sup>Terrestrial Ecology Unit, Ghent University, Ghent, Belgium

<sup>3</sup>German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Leipzig, Germany

<sup>4</sup>Department of Evolutionary Theory, Max Planck Institute for Evolutionary Biology, Plön, Germany

<sup>5</sup>Institute for Biology and Environmental Sciences, Carl von Ossietzky University, Oldenburg, Germany

<sup>6</sup>Department of Life Sciences, Imperial College London, Ascot, United Kingdom

## Correspondence

Adriana Alzate, Groningen Institute for Evolutionary Life Sciences, University of Groningen, Box 11103, 9700 CC Groningen, The Netherlands.  
Email: adria.alzate@gmail.com

## Funding information

Nederlandse Organisatie voor Wetenschappelijk Onderzoek, Grant/Award Number: 865.13.003; Fonds Wetenschappelijk Onderzoek, Grant/Award Number: G.018017.N.; Ubbo Emmius Fund; BelSpo IAP; Natural Environment Research Council, Grant/Award Number: NE/I021179 and NE/L011611/1

Editor: Jonathan Belmaker

## Abstract

**Aim:** The great variation in range sizes among species has fascinated ecologists for decades. Reef-associated fish species live in highly spatially structured habitats and adopt a wide range of dispersal strategies. We consequently expect species with greater dispersal ability to occupy larger ranges. However, empirical evidence for such a positive relationship between dispersal and range size remains scarce. Here, we unveil the role of dispersal on the range size distribution of reef-associated fishes using empirical data and a novel spatially explicit model.

**Location:** Tropical Eastern Pacific.

**Major taxa studied:** Reef-associated fishes.

**Time period:** Underlying records are from the 20th and 21st centuries.

**Methods:** We estimated range size distributions for all reef-associated fishes separated into six guilds, each with different dispersal abilities. We used a one-dimensional spatially explicit neutral model, which simulates the distribution of species along a linear and contiguous coastline, to explore the effect of dispersal, speciation and sampling on the distribution of range sizes. Our model incorporates biologically important long-distance dispersal events with a fat-tailed dispersal kernel and also adopts a more realistic gradual “protracted” speciation process than originally used in neutral theory. We fitted the model to the empirical data using an approximate Bayesian computation approach, with a sequential Monte Carlo algorithm.

**Results:** Stochastic birth, death, speciation and dispersal events alone can accurately explain empirical range size distributions for six different guilds of tropical, reef-associated fishes. Variation in range size distributions among guilds are explained purely by differences in dispersal ability with the best dispersers being distributed over larger ranges.

**Main conclusions:** Neutral processes and guild-specific dispersal ability provide a general explanation for both within- and across-guild range size variation. Our results

\*Joint last authors

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2019 The Authors Global Ecology and Biogeography Published by John Wiley & Sons Ltd

support the theoretically expected, but empirically much debated, hypothesis that high dispersal capacity promotes the establishment of large range size.

#### KEYWORDS

dispersal, neutral model, range size, range size distribution, reef fishes, spatially explicit

## 1 | INTRODUCTION

What is driving the large natural variation in the range size of species (Gaston, 2003)? Ultimately, only a few ecological processes should be important in determining a species' range size: dispersal to a new habitat, successful colonization of that habitat and (avoidance of) local extinction (Brown & Kondric-Brown, 1977; Gaston & He, 2002; Hanski, 1982; Holt & Gomulkiewicz, 1996; MacArthur & Wilson, 1967). Besides these ecological processes, macroevolutionary processes such as speciation and extinction also have important implications for range sizes as they affect the creation, division and removal of ranges at larger temporal scales (Anderson, 1985; Gaston & Chown, 1999). Dispersal is central to all the above processes, because it is key both for the colonization of new habitats and for population persistence in suboptimal habitats, which may otherwise go extinct (Brown & Kondric-Brown, 1977; MacArthur & Wilson, 1967). Furthermore, dispersal promotes gene flow, which can impede speciation while also bringing the genetic variability necessary for adaptation and successful range expansion (Holt & Gomulkiewicz, 1996).

One group of organisms for which dispersal seems to be an especially important driver of range size is reef fishes. These organisms occupy habitats that are by nature highly fragmented. The ability to disperse to these habitats should therefore be important for colonization, establishment and range expansion. However, despite theoretical expectations predicting a positive association between dispersal and range size (reviewed in Lester, Ruttenberg, Gaines, & Kinlan, 2007), empirical evidence for this in reef fishes remains scarce (Lester & Ruttenberg, 2005; Luiz et al., 2013; Mora et al., 2012; Ruttenberg & Lester, 2015). There are many possible explanations for the apparent lack of a positive range size–dispersal relationship; these reflect the many processes that potentially drive range size including speciation, local extinction, and range dynamics or changes during a species' lifetime (Webb & Gaston, 2000, reviewed in Gaston, 2003). Firstly, range size is likely to vary with species age: older species might attain larger ranges than newly formed species (Webb & Gaston, 2000). However, until complete information on the age of all species is available, it will be difficult to test this and correct for it. Secondly, species range dynamics are affected by numerous ecological and evolutionary factors including biological interactions and the species' behavioural and functional traits (Stahl, Reu, & Wirth, 2014). Thirdly, sampling intensity and detection probability vary across space and across species (Alzate, Zapata, & Giraldo, 2014; Dennis, Sparks, & Hardy, 1999), and such sampling biases may also drive variation in the apparent range size. Finally, stochastic events, especially during the early life of a species,

may bring additional noise to the present-day range size, making it difficult to find general patterns.

Dispersal is a complex trait that depends on an individual's life stage in ways that are not easily quantifiable, for example, during the departure (initiation of dispersal, for instance during spawning), transfer (the pelagic stage of eggs and larvae) and settlement phases (Bonte et al., 2012). Due to this complexity, direct measurement of the entire dispersal process is hard to achieve; instead, researchers have used proxies of dispersal: traits that, based on the literature, are linked to dispersal (e.g., in fish body size, pelagic larval stage, type of egg). The choice of traits to investigate and the way to measure them may influence the outcome of studies examining the role of dispersal on range sizes. For example, many correlative studies of reef fish dispersal have focused primarily on the larval stage (Lester & Ruttenberg, 2005; Lester et al., 2007; Mora et al., 2012, but see Luiz et al., 2013), thereby neglecting evidence that dispersal also occurs in earlier life stages as eggs and in later life stages as adult fishes (Addis, Patterson, Dance, & Ingram, 2013; Alzate, van der Plas, Zapata, Bonte, & Etienne, 2019; Appeldoorn, Hensley, Shapiro, Kioroglou, & Sanderson, 1994; Kaunda-Arara & Rose, 2004; Leis, 1978).

Pattern-orientated correlative studies, which test for associations between traits related to dispersal and range size, fail to incorporate more mechanistic insights (Connolly, Keith, Colwell, & Rahbek, 2017). In contrast, mechanistic models make testable predictions by explicitly accounting for the more fundamental processes affecting range size. Previous mechanistic studies have attempted to explain range size using colonization–extinction models (Hanski, 1982) or models of population dynamics (Gaston & He, 2002). However, they were not developed to explain variation in range size across many species exploring several mechanisms. Here, we apply a variation of the unified neutral theory of biodiversity and biogeography (Hubbell, 2001), originally used to explain other macroecological patterns such as species abundance distributions, species–area relationships and beta diversity. One previous study used a neutral model to reproduce both spatial patterns in species richness and range size distributions, but it focused on short time spans rather than equilibrium behaviour and as such did not incorporate speciation (Rangel & Diniz-Filho, 2005). Here, we extend the neutral model of Hubbell (2001) to include spatially explicit dynamics and a more realistic speciation process (Rosindell, Cornell, Hubbell, & Etienne, 2010; Rosindell, Wong, & Etienne, 2008), both of which we expect to be important for a study of interspecific variation in range sizes. This mechanistic model provides a way to quantitatively assess how dispersal can influence species range size distributions, while at the

same time considering other interacting mechanisms, including sampling and speciation, which are known to affect range size (Gaston, 2003). We test the ability of our model to explain variation in range size by comparing its predictions against empirical range size distributions of a complete reef fish assemblage in a well-defined region: the Tropical Eastern Pacific (TEP). We make predictions of range size distributions for each of six distinct guilds with different dispersal characteristics. Dispersal is based on mobility traits (not directly measured) acting in both the early (egg and larval) and later (adult) life stages. Our model is neutral within each guild and so excludes any within-guild niche-based processes and individual differences. Crucially, by applying independent neutral models to each of the six guilds, we are able to focus on studying the effects of different dispersal abilities for each guild in isolation from other complicating factors such as environmental preference. With our spatially explicit model, we test firstly whether range size distributions within guilds of reef fishes can be explained by neutral factors alone, and secondly whether variation in range size distribution across guilds can be explained by differences in dispersal ability. Our spatially explicit neutral model, which incorporates stochastic birth–death dynamics, speciation and dispersal, can generate testable predictions on range size distribution and species richness patterns.

## 2 | METHODS

### 2.1 | Reef-associated fish data

From the online database “Shorefishes of the Tropical Eastern Pacific (SFTEP)” (Robertson & Allen, 2016), we collated spatial coordinates of species occurrences (43,810 records) for all bony fishes (575 species) associated with reef habitats reported in the TEP. We used only records inside the TEP region: between 24° N (outer coast of the Gulf of California, including all the inner coast) and 4° S (SFTEP, Robertson & Allen, 2016).

Reef fish species were classified into six different dispersal guilds according to traits related to dispersal: spawning mode and adult mobility. We classified spawning mode into two types: pelagic and non-pelagic. Differences in this early life history trait might confer diverse capacities for dispersal (Leis et al., 2013; Riginos, Douglas, Jin, Shanahan, & Tremblay, 2011). Pelagic spawners release their eggs in the water column, which are passively transported by water currents until the larvae hatch and are able to better control active swimming (Leis et al., 2013; Stobutzki & Bellwood 1997). This increase in pre-hatching dispersal might have strong effects on overall dispersal in the pelagic environment (Leis et al., 2013). Contrary to pelagic spawners, for which both the egg and larval phases are pelagic, non-pelagic spawners either attach their eggs to the substrate, are live-bearers, or keep their eggs in the mouth or pouch until they hatch. Their larvae usually emerge at larger sizes and are more mature than the larvae of non-pelagic spawners (Leis et al., 2013; Wootton, 1992), resulting in an early control of active swimming, therefore affecting dispersal by reducing the pelagic interval (Bonhomme & Planes, 2000; Burgess, Baskett, Grosberg, Morgan, & Strathmann,

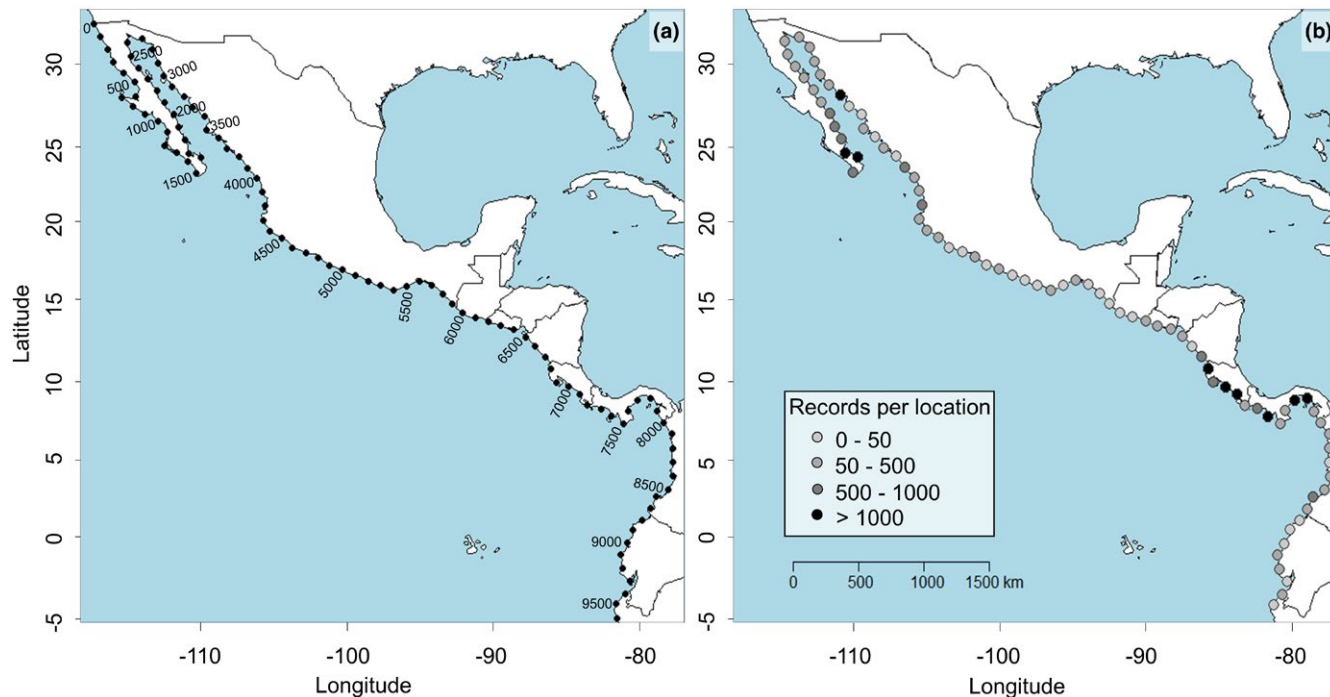
2015; Leis, 2006; Leis et al., 2013; Munday & Jones, 1998). We classified adult mobility following Floeter, Ferreira, Dominici-Arosemena, and Zalmon (2004) as low, medium and high. Low adult mobility is associated with site-attached species occupying small home ranges ( $< 10 \text{ m}^2$ ). Medium adult mobility denotes species that are weakly mobile, relatively sedentary, with close association to the substrate and that can be distributed over the entire reef area ( $< 1,000 \text{ m}^2$ ). High adult mobility is represented by species that show a wide horizontal displacement and occur in the water column (Floeter et al., 2004). Mobility for each species was assigned depending on the taxonomic level at which information was reported: species, genus or family adult mobility. In some cases, mobility information was not available, but could be assigned according to the biology of the species, for example, pearlfishes (family Carapidae), which are known to live inside the anal pore of sea cucumbers, were all classified as having low adult mobility. Information on spawning mode was obtained from the SFTEP online database (Robertson & Allen, 2016). Pelagic larval duration, although often used when studying range size of reef fishes, is not known for the majority (69%) of species in the TEP region, so we cannot use it for this study.

### 2.2 | Measuring range size

The range size of each species was calculated using a novel metric developed for maximizing comparability between simulated and observed range sizes: coastline distance. In contrast with other traditional metrics, for example, maximum linear distance and latitudinal and longitudinal extent (Gaston, 1996), coastline distance does not underestimate or overestimate range size due to the particular spatial configuration of the TEP (Supporting Information Appendix S1). We defined coastline distance as the contour distance (measured using units of 100 km) between the most distant points along the coastline where the species was reported. The east and west coasts of the Gulf of California are collapsed into a single line of habitat because the distance between opposite sides of the gulf is likely too small to substantially restrict dispersal across the gulf. All distance measurements were calculated in kilometres using the function `geo-dist` from the R package `gmt` (Magnusson, 2015) and transformed into relative values, where 100% is the coastline distance between the latitudes 24° N and 4° S.

### 2.3 | Richness gradients

To examine the species richness gradient along the TEP coast, we calculated species richness for each segment of 100 km of coastline using all species within the six dispersal guilds. In addition, we calculated richness only for the TEP endemic species within the six dispersal guilds and for the TEP non-endemic species. A species is considered to occur in each location (shown in Figure 1a) within its geographical range. For consistency, we treated location in the same way as the estimation of range sizes: east and west coasts of the Gulf of California were collapsed into a single line of habitat, so locations occurring at opposite sides of the Gulf were merged.



**FIGURE 1** (a) Map with “locations” every 100 km along the Tropical Eastern Pacific (TEP) coastline. For all analyses, we used a “collapsed” Gulf of California and excluded the outer part of the Peninsula of California. (b) Map showing number of records per location along the coastline

The outer part of the Peninsula of California was excluded from analyses (Figure 1b).

Our coastline distance metric treats the coast of the TEP as one-dimensional (distance is measured around the coast, not in a direct line along the shortest distance); this maximizes the comparability of empirically observed range sizes with those simulated by our one-dimensional model. We excluded observations from oceanic islands when quantifying range sizes, again to maximize comparability with simulated ranges.

## 2.4 | Null model

We used a null model (excluding spatial autocorrelation) to test whether the observed range size distributions can be obtained without dispersal limitation. For this we used the data on spatial (xy) coordinates and species IDs for each dispersal guild in the TEP. Firstly, we stored two vectors: one with xy coordinates and other one with species IDs. Then we randomly permuted the position of the species IDs along the vector and assigned each species ID to a new xy coordinate. Range sizes were measured as described above for the empirical data and range size distributions were obtained for the six different dispersal guilds. We simulated 100 of these null range size distributions by different permutations of the species IDs.

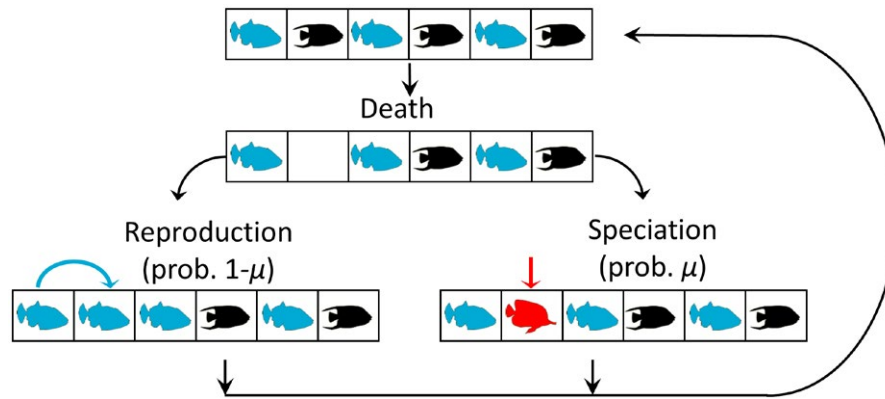
## 2.5 | Spatially explicit neutral model

We used a one-dimensional spatially explicit neutral model to simulate the spatial distribution of species along a linear

coastline. This configuration best reflects the particular geographical distribution of reefs (coral and rocky) in the TEP region: a long coastline with a narrow continental platform. As in the original neutral model (Hubbell, 2001), the habitat is saturated (zero-sum dynamics), and the species identity of an individual has no bearing on its chances of dispersal, mortality, reproduction, the initiation of speciation or the completion of speciation. The dynamics proceed as follows: at every time step one individual, chosen at random according to a uniform distribution, dies and is replaced by a new incipient species (with a probability  $\mu$ ) or by the newborn offspring of an existing individual (with a probability  $1 - \mu$ ) (see Figure 2 for a schematic representation of the model). The position  $X$  of the parent of the new offspring is determined by a Pareto dispersal kernel, which describes long-distance dispersal well, in line with empirical dispersal distributions of reef fishes (Jones 2015):

$$f(X) = \begin{cases} \frac{\alpha X_m^\alpha}{X^{\alpha+1}}, & X \geq X_m \\ 0, & X < X_m \end{cases} \quad (1)$$

where  $X_m$  is a scale parameter (the minimum dispersal distance), and  $\alpha$  is a shape parameter that changes the distribution from an exponential-like distribution (large value of  $\alpha$ ) to a fat-tailed distribution (lower values of  $\alpha$ ) where many short-distance dispersal events are combined with an occasional extreme long-distance dispersal event. Random samples from the distribution can be calculated using the inverse random sampling formula for the range size  $T$ :



**FIGURE 2** Conceptual representation of the spatially explicit neutral model algorithm

**Notes.** At every time step one random individual along the lattice dies (in this case a black fish) and is replaced either by a new incipient species (red fish), which may become a good species if it survives for sufficiently long, or by the newborn offspring of an existing individual (blue fish).

$$T = \frac{X_m}{U^{1/\alpha}} \quad (2)$$

where  $U$  is a random variable drawn from a uniform distribution between 0 and 1. To separate the effects of the shape of the distribution and the mean dispersal distance ( $X_{\text{mean}}$ ), we rescaled the inverse random sampling formula for the Pareto distribution so that it is written in terms of the mean dispersal distance  $X_{\text{mean}}$ :

$$T = \frac{\frac{\alpha-1}{\alpha} X_{\text{mean}}}{U^{1/\alpha}} \quad (3)$$

In contrast to the typical ecological neutral model, we assumed that speciation is a gradual “protracted” process rather than an instantaneous event (Etienne & Rosindell, 2012; Rosindell et al., 2010). When a birth event takes place, an incipient species can form with probability  $\mu$ ; the newborn is still considered conspecific to its parent, but if sufficient time passes and descendants of the newborn individual survive, these descendants will be considered a new good species. This protracted speciation model entails one extra parameter  $\tau$ : “protractedness”, the number of generations required for an incipient species to become a good species. One generation corresponds to half of a complete turnover of the community because generations overlap. Both speciation probability and protractedness influence the generation of new species; the true speciation rate is a function of both parameters ( $\mu/(1 + \tau)$ ) as described by Rosindell et al. (2010). We simulated the spatially explicit neutral model using a coalescence approach (Rosindell et al., 2008), which improves simulation efficiency, while guaranteeing the metacommunity is sampled at dynamic equilibrium and thus eliminating the problem of determining an appropriate “burn-in time” for the simulations.

Our simulation model was written in C++ and all post-simulation analyses were performed with R, version 3.3.1 (R Core Team, 2016).

## 2.6 | Model behaviour

We explored the effect of dispersal on the distribution of range sizes by running simulations using various dispersal kernels, which

differ in their  $X_{\text{mean}}$  and  $\alpha$  parameter values. We used a linear lattice composed of 50,000 “units”, which can be thought of as individual organisms or larger cohorts of individuals behaving in a similar manner (Harfoot et al., 2014). We found that larger lattices produce similar results (Supporting Information Appendix S2) but are computationally intractable for parameter fitting exercises that require many successive simulation runs. As in the real world not all individuals are sampled, the proportion of sampled individuals (sampling percentage) could therefore affect the observed distribution of ranges. Sampling was performed by randomly choosing individuals along the linear lattice, and only sampled individuals were used to quantify range sizes, thus taking into account the effect of sampling on apparent range. Although sample areas along the TEP are not random, sampling in a realistic manner produces virtually identical results to random sampling (Supporting Information Appendix S3). In the real world, suitable habitat is often not contiguous; the same is true for the TEP, where reefs are fragmented by long stretches of sand or other soft substrates (Supporting Information Appendix S4). We explored the effect that habitat fragmentation has on the geographical distribution of species, by adding into the model areas of suitable and unsuitable habitat. We did so by introducing sites within the lattice that were designated as unsuitable habitat. This introduces barriers to dispersal where many such contiguous sites are located. The locations of the unsuitable sites were based on the real distribution of reefs in the TEP (Supporting Information Appendix S4), scaled to our lattice size.

We examined the effect of dispersal ( $X_{\text{mean}}$  and  $\alpha$ ), speciation initiation rate ( $\mu$ ), speciation protractedness ( $\tau$ ) and sampling percentage ( $s$ ) on the distribution of species’ range sizes. As species age is suggested to be positively correlated to range size (Gaston, 2003), we also explored the effect of interspecific variation in speciation rates on the distribution of range sizes. When speciation rate is high, species are on average younger, which may have an effect on range size. In addition, we explore how the configuration of the habitat (fragmentation) affects the distribution of species ranges.



In our default scenario, we used a lattice with contiguous habitat and the following parameter values: mean dispersal distance  $X_{\text{mean}} = 2\%$ , dispersal kernel shape  $\alpha = 3.0$ , sampling percentage  $s = 100\%$ , speciation probability  $\mu = 0.0005$  and protractedness  $\tau = 10$ . We simulated five sets of alternative scenarios, in which values of either  $X_{\text{mean}}$ ,  $\alpha$ ,  $s$ ,  $\mu$  or  $\tau$  were altered as well as a scenario showing the effects of habitat fragmentation for high and low dispersers. In particular, we explored five different mean dispersal distances ( $X_{\text{mean}} = [2, 5, 10, 20, 40\%]$ ) on a lattice with either a contiguous or a fragmented habitat, five different  $\alpha$ -values ( $\alpha = [1.5, 2.0, 2.5, 3.0, 3.5]$ ), five different sampling percentages ( $s = [1, 5, 20, 50, 100\% \text{ of all individuals}]$ ) and four different speciation probability and protractedness values ( $\mu = [5 \times 10^{-2}, 5 \times 10^{-3}, 5 \times 10^{-4}, 5 \times 10^{-5}]$ ,  $\tau = [0, 10, 100, 1000]$ ).

We estimated the simulated range size for each species as the linear distance (equivalent to coastline distance in the one-dimensional lattice) between the most distant points where the species is recorded. The range size was measured relative to the total lattice size and transformed to percentages (100% corresponding to the total size of the complete lattice). We replicated the simulations 100 times and calculated mean and 95% Confidence intervals (CI) values.

## 2.7 | Model fitting

To estimate parameters from the artificial range size distributions, we used an approximate Bayesian computation approach, with a sequential Monte Carlo algorithm (ABC-SMC) as described by Toni, Welch, Strelkova, Ipsen, and Stumpf (2009). To assess the similarity between the artificial data and simulation outcomes, we calculated the sum of squares between the inverse cumulative distribution for the simulated and artificial data, based on the differences in both the range size distributions and species richness levels. Progression of the acceptance threshold was modelled as an exponentially decreasing function, where the threshold at iteration  $t$  of the ABC-SMC algorithm was  $500e^{-\frac{1}{4}t}$ . We assumed the following prior distributions for each parameter (on a  $\log_{10}$  scale, e.g.,  $U_{10}(0,1) = 10^{U(0,1)}$ , where  $U$  is a uniform distribution),  $X_{\text{mean}}$ :  $U_{10}(-4, -0.25)$ ,  $\alpha$ :  $U_{10}(0,1)$ , speciation initiation rate:  $U_{10}(-4, 0)$ , protractedness:  $U_{10}(0,5)$  and sampling:  $U_{10}(-4, 0)$ . Per ABC-SMC iteration, we used 10,000 particles. The ABC-SMC algorithm ran for 20 iterations, or until the acceptance rate dropped below 1 in 1,000,000 proposed parameter combinations. Perturbation of the parameters was performed on a  $\log_{10}$  scale, to avoid parameters reaching a negative value. Parameters were perturbed by first taking the  $\log_{10}$ , then adding a random number drawn from a normal distribution with mean zero and standard deviation 0.05, after which we exponentiated the parameter again. Finally, we checked whether the parameter values still lay within the prior ranges; if not, the particle was rejected. For each dataset, we performed 10 replicate fits. Posterior parameter values were corrected using linear regression (Beaumont, Zhang, & Balding, 2002).

To assess the impact of speciation, we performed two additional fits using a modified speciation model. Firstly, speciation rate was fixed at 1, reflecting a fixed age for all species, which is given by the protractedness parameter. Secondly, we used a fixed protractedness

of zero, reflecting a model where speciation is instantaneous (point mutation). However, posterior fits of these models [corrected for a reduced number of parameters, following the Akaike's information criterion (AIC; Akaike, 1974)] show that they fit the data poorly in comparison to the full model (Supporting Information Appendix S5), and they were thus not considered further.

Lastly, goodness of fit of the model to the data was estimated by calculating the posterior predictive  $p$ -value, using 1,000 replicates in the function "gfit" in the "abc" package in R (Csilléry, François, & Blum, 2012).

## 2.8 | Model fitting validation

Prior to fitting the model to the empirical data, we assessed the accuracy of our inference method. For this, we generated artificial datasets using known parameters for  $X_{\text{mean}}$ ,  $\alpha$ , speciation, sampling and protractedness. In particular, we used  $X_{\text{mean}} = 0.001, 0.01, 0.1$  or  $0.2$ ,  $\alpha = 2, 4, 6$  or  $8$ ,  $s = 0.025$  or  $0.25$ , and two different speciation regimes: one with high speciation ( $\mu = 0.01$ ) and high protractedness ( $\tau = 2,500$ ), and one with low speciation ( $\mu = 0.001$ ) and low protractedness ( $\tau = 25$ ). For each parameter combination, we generated 10 artificial datasets. We then attempted to recover the known simulated parameters from simulated data using the ABC-SMC approach. In total, we performed  $(10 \times 4 \times 4 \times 2 \times 2) = 640$  ABC-SMC inferences to assess accuracy. Posterior distributions of parameter values generally closely matched the simulated parameter values (Supporting Information Appendix S6), which indicates that our fitting procedure is appropriate for estimating the parameter values of our neutral model. Only in the case of  $\alpha$  (dispersal kernel shape) were the parameter estimates not very accurate. This was likely due to the low power of  $\alpha$  in explaining range size variation.

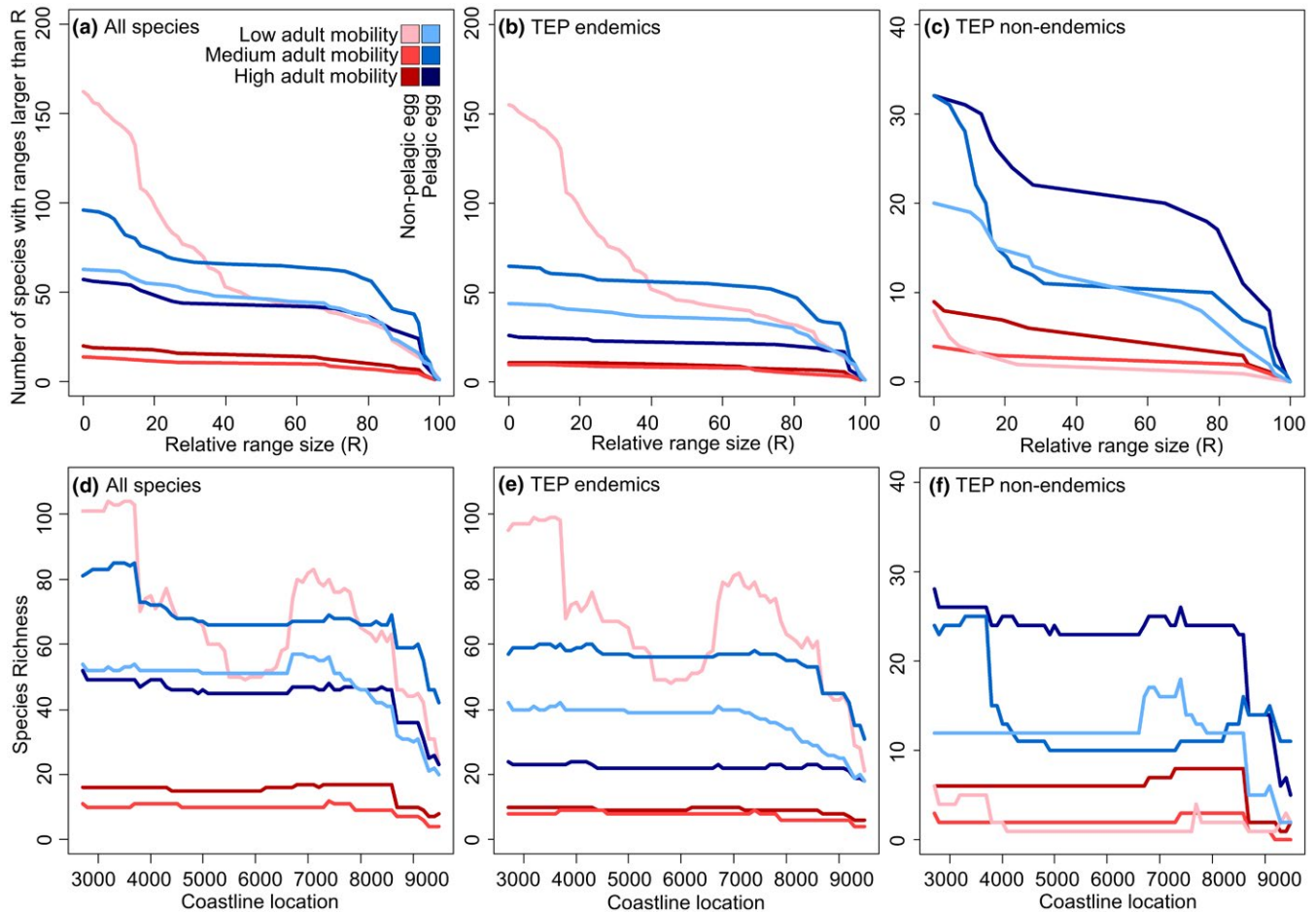
## 2.9 | Theoretical richness gradients

In order to make predictions of species richness gradients along the TEP coastline, we first fitted our model to the empirical range size distribution using a habitat fragmented in the same pattern as the real-world habitat (Supporting Information Appendix S4). Next, we used the fitted parameters for each dispersal guild to run 100 replicate simulations. Finally, we quantified species richness along the complete linear lattice and plotted the results alongside empirical richness gradients for comparison.

# 3 | RESULTS

## 3.1 | Empirical range size distributions

Pelagic spawners have a relatively high proportion of species with large ranges, irrespective of their adult mobility (Figure 3a). Specifically, more than half of the species have ranges larger than 80% of the maximum possible range for the studied region. In contrast, the range size distribution of non-pelagic spawners strongly depends on the capacity of adult fishes to disperse. Within the non-pelagic spawners, the lowest dispersive guild has the highest



**FIGURE 3** Distribution of range sizes and richness patterns along the Tropical Eastern Pacific (TEP) coast for different dispersal guilds of reef fishes in the TEP Notes. A guild is defined as a group of species that share the same spawning mode (pelagic and non-pelagic spawners) and adult mobility (low, medium and high). Range size is shown in relative terms, where a range of 100% is a species that covers the entire region. We used coastline distance as the range size metric, which is the distance between the most distant points along the coastline. Individuals from oceanic islands are excluded to be consistent with the one-dimensional nature of the model. The two sides of the Gulf of California coastline were collapsed into a single line of habitat. The distribution of ranges for each guild is shown as a cumulative distribution curve, which shows the proportion of species (y axis) with a range larger than a given size (x axis). Range size distributions are shown for (a) all species in the TEP, (b) all TEP endemic species and (c) all TEP non-endemic species. Richness gradients along the TEP coast are shown for (d) all species in the TEP, (e) all TEP endemic species and (f) all TEP non-endemic species.

number of species, the highest proportion of species with small ranges and the lowest proportion of species with large ranges (Figure 1a). While more than half of the non-pelagic spawning species with medium or high adult mobility have ranges larger than 80% of the maximum range, for species with low mobility, only a fifth of species have ranges larger than 80% of the maximum. This general pattern is even more pronounced for endemic species: a large majority of the pelagic spawners and non-pelagic spawners with medium and high adult mobility have large ranges, and non-pelagic spawners with low adult mobility have an even higher proportion of species with small ranges (Figure 3b). In contrast, most of the non-endemic species are pelagic spawners, and for these species, the range size distribution is bimodal (especially for medium mobile species): while there are a large number of species with small ranges, there are others displaying large ranges and a

few showing intermediate ranges (Figure 3c). There are relatively few non-endemic non-pelagic spawning species. Those with medium or high mobility tend to have large ranges, while the majority of non-pelagic non-endemic species with small ranges have low adult mobility (Figure 3c).

### 3.2 | Empirical richness gradients

Species richness in general remains relatively constant for all guilds (except for non-pelagic spawners with low adult mobility) along the TEP coast, decreasing in the last part of the curve (Figure 3d). However, some curves (e.g., pelagic spawners with low and medium adult mobility) show peaks at locations 2,600–3,500 and around location 7,000 (see Figure 1a for a map of the locations). Richness for non-pelagic spawners with low adult mobility is maximum at the

first part of the curve, showing an abrupt drop at location 3,600, then constantly decreasing until reaching a valley between locations 5,500 and 7,000. After this, richness increases to 80 species, followed by a constant decrease (Figure 3d). When examining only endemic species (Figure 3e), we see that all guilds except non-pelagic and low adult mobility guilds show a constant richness, without richness peaks. For non-endemic species, species richness is in general lower than for endemic species (Figure 3e), and non-endemic pelagic spawners attain a higher richness than non-endemic non-pelagic spawners along the TEP coast. Pelagic spawners with high adult mobility have a constant richness along the coast that decreases at higher locations (towards the southern edge of the TEP). Non-endemic pelagic spawners with medium adult mobility, however, show a bimodal pattern, with one peak at locations 2,600–3,600 and another at 8,200–9,000. Non-endemic pelagic spawners with low adult mobility show a peak at locations 6,600–7,500.

### 3.3 | Null model

In the absence of dispersal limitation, our null model predicts species to be more widespread than they are in reality. Therefore, the null model cannot explain the empirical range size distribution of reef fishes in the TEP and especially not for the less dispersive guilds. All dispersal guilds are predicted under the null model to have a larger proportion of species with large range sizes than the observed ones (Figure 4).

### 3.4 | Spatially explicit neutral model

The strongest effects on the distribution of range sizes are caused by variation in mean dispersal distance ( $X_{\text{mean}}$ ), speciation rate and protractedness (Figure 5). Although dispersal ( $X_{\text{mean}}$  and  $\alpha$ ) has a strong effect on the shape of the range size distribution, the contributions of  $X_{\text{mean}}$  and  $\alpha$  to the effect of dispersal on the range size distribution are not equal, with the majority of the dispersal effect resulting from  $X_{\text{mean}}$  (Figure 5a,b). As  $X_{\text{mean}}$  increases, the proportion of species with large ranges increases as well. The shape parameter of the dispersal kernel ( $\alpha$ ) has limited influence on the distribution of range sizes (Figure 5b). Speciation exerts a strong effect on the distribution of ranges, with a higher proportion of species having a large range size when speciation rate is low (Figure 5c). A high speciation rate produces more new species, which initially have small ranges, and hence a decrease in the number of species with large ranges, and a (potentially unrealistically) high number of species in total (Figure 5c). The effect of protractedness is similar to that of speciation, as they both modify the number of species and the rate at which these are created. The higher the protractedness, the longer the time before an incipient species becomes a good species, and as a result fewer (good) species have small ranges (Figure 5d). Habitat fragmentation affects the range size distribution of low dispersive species more strongly than of high dispersive species (Figure 5e). Sampling affects the distribution of ranges differently from dispersal, speciation or protractedness: a lower sampling effort leads to

more species with few individuals and thus a higher proportion of species with apparently small ranges (Figure 5f).

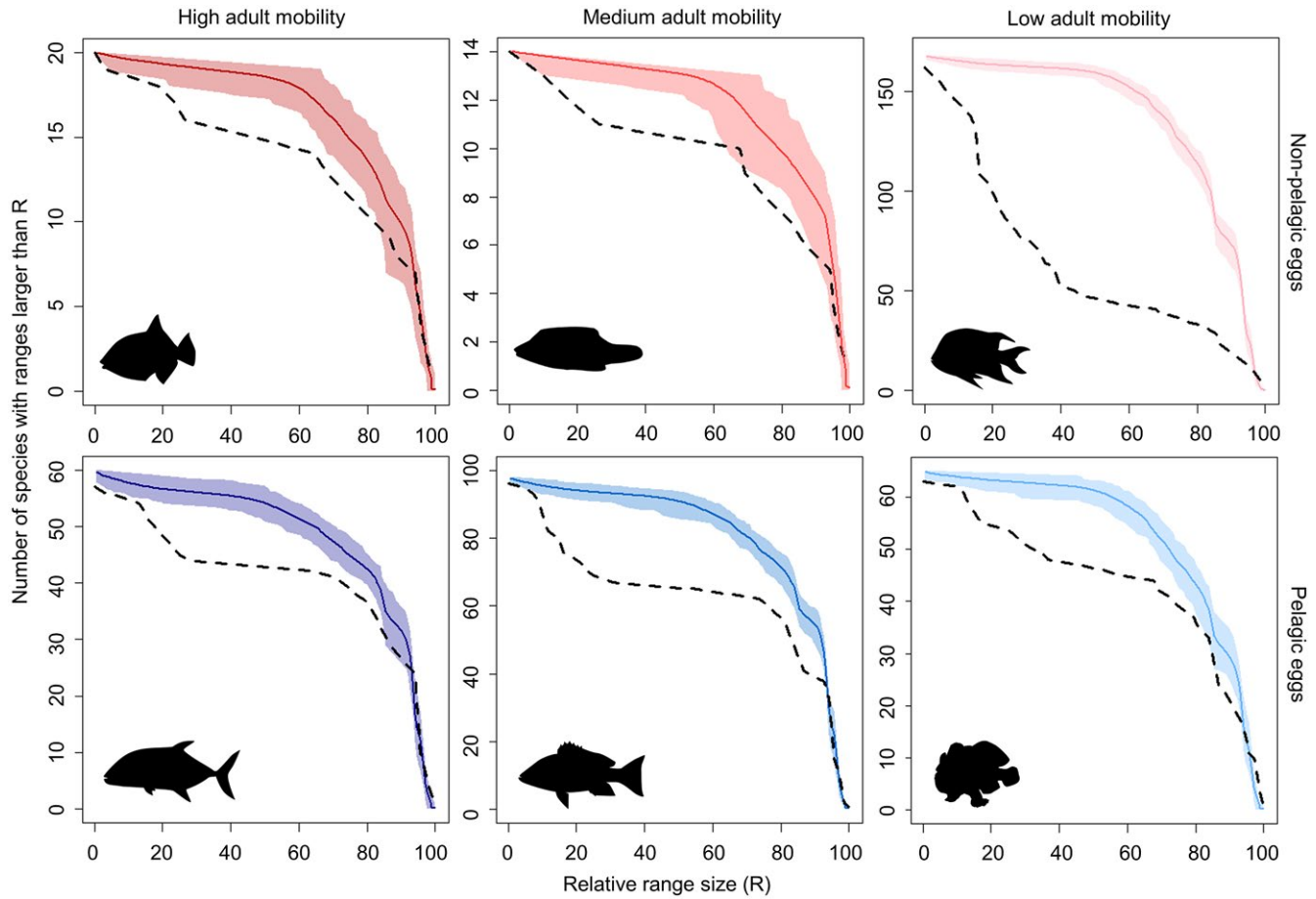
### 3.5 | Model fitting

The fitting procedure on the empirical range size distributions for the six dispersal guilds of reef fishes showed a good fit for all datasets (posterior predictive  $p$ -value  $> 0.05$ , Supporting Information Appendix S7a), except for non-pelagic low mobility and for pelagic high mobility ( $p = 0.012$  and  $0.027$  respectively, Supporting Information Appendix S7a, Figure 6). However, posterior predictive testing is confounded here by differences in variation across simulations across datasets. Comparing the normalized sum of least squares (normalized for the total number of species in the dataset), we find that for the non-pelagic low mobility and the pelagic high mobility guilds, the average normalized sum of least squares is lower than for the other datasets (Supporting Information Appendix S7c). This means that even though the model cannot exactly reproduce the empirical pattern (as suggested by the posterior predictive  $p$ -values, Supporting Information Appendix S7a), the model predictions are nonetheless really close to the empirical pattern (indicated by the low values of the sums of least squares, Supporting Information Appendix S7b,c). Fitting the neutral model with habitat fragmentation to the empirical range size distributions produced similar results to the ones without fragmentation (Supporting Information Appendix S8). However, for the majority of dispersal guilds, the model without habitat fragmentation showed a better fit than the model with habitat fragmentation (Supporting Information Appendix S7), it yielded higher sums of least squares values and lower posterior predictive  $p$ -values, indicating a worse fit to the data (Supporting Information Appendix S7).

In line with expectations, estimated mean dispersal distances for each guild were largest for the guilds with the highest proportion of large ranges: pelagic spawners and high adult mobility. The  $\alpha$ -values were similar for all dispersal guilds (between 2.7 and 4.23). Estimated sampling completeness was lowest for the guilds of non-pelagic spawners with high and medium mobility (0.7 and 0.3%, respectively), similarly low for the guild of pelagic spawners (3%–8%) and highest for the guild of non-pelagic spawners with low adult mobility (33%). Protractedness values (time taken for an incipient species to become a good species) were lowest for non-pelagic low mobility species, intermediate for pelagic spawners and highest for non-pelagic spawners with high and intermediate mobility. The speciation rate (per generation probability for an individual to become a new incipient species) was lowest for pelagic spawners ( $< 0.0001$ ) and highest for non-pelagic spawners (0.001–0.002). See Supporting Information Appendix S9 for a complete description of the model estimates and Figure 7 for the distribution of posteriors.

Although the posterior predictive  $p$ -values were only above 0.05 for two dispersal guilds (pelagic spawners with high and medium adult mobility), inspection of the qualitative properties of the fittings shows bimodal behaviour in all guilds that the neutral model could not fully replicate within its parameter space, for example, for pelagic spawners with medium adult mobility (Figure 6). We explain this





**FIGURE 4** Range size distributions predicted by the null model without dispersal limitation and geographical constraints

Notes. Coloured bands represent the model outcomes over 100 replicates [solid lines: mean; shaded area: 95% Confidence intervals (CI)]. Dashed lines represent the empirical data.

bimodality in the overall distribution as the result of combining the distribution of TEP endemic and non-endemic species (Figure 1a–c).

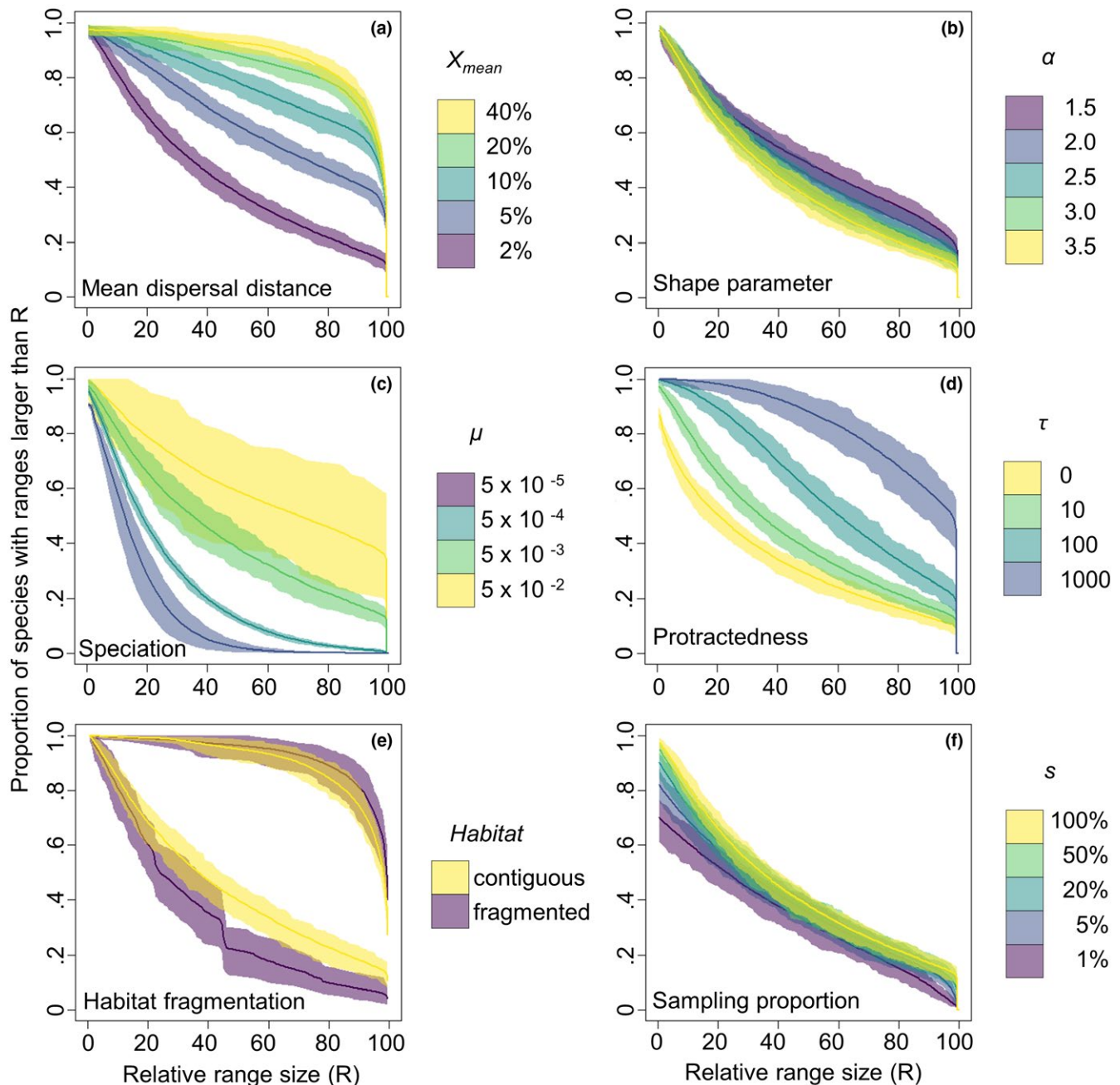
### 3.6 | Theoretical richness gradients

Comparing the richness gradients predicted by the model fitted to the range size distribution with habitat fragmentation to the empirical richness gradients along the TEP coast, we show that our model supports qualitatively the basic pattern of species richness gradients in the TEP, especially for species appearing outside the first 18% of the region (Figure 8). Deviations from the empirical data for the first 18% of the region are likely caused by collapsing the two coasts of the Gulf of California in the empirical data. While this artefact does not have a strong effect on the range size distributions (Supporting Information Appendix S1), it does affect richness patterns. Specifically, while for pelagic fishes collapsing the Gulf of California was of no consequence, it was important for the low mobility non-pelagic spawners (and even a little for the medium mobility non-pelagic spawners). Furthermore, the enclosed nature of the Gulf of California might isolate that area and thus promote speciation (which is not explicitly considered in our model). For pelagic spawners, the model predicts richness to increase rapidly until

reaching a plateau in the centre of the distribution, followed by a more gradual decrease in the last part of the distribution in particular for low and medium adult mobility species. For non-pelagic spawners with medium and high adult mobility, the model predicts a similar pattern as for pelagic spawners with medium and low adult mobility (Figure 8). The model, however, predicts a very different pattern for species with non-pelagic eggs and low adult mobility: three richness peaks corresponding to areas where reef habitat is available.

## 4 | DISCUSSION

Macroecologists have long sought to understand the large variation in range sizes across species by correlating species ranges with other factors, such as traits, that are believed to influence range size. Despite the usefulness of this approach for hypothesis testing, such correlational studies are not able to provide a deeper understanding about the mechanisms behind the macroecological patterns we see in nature. In order to better understand the underlying mechanisms, we developed a model that can capture the way mechanisms act together to produce biodiversity patterns. Models can vary in complexity and in the predictions they provide. For instance, simple null

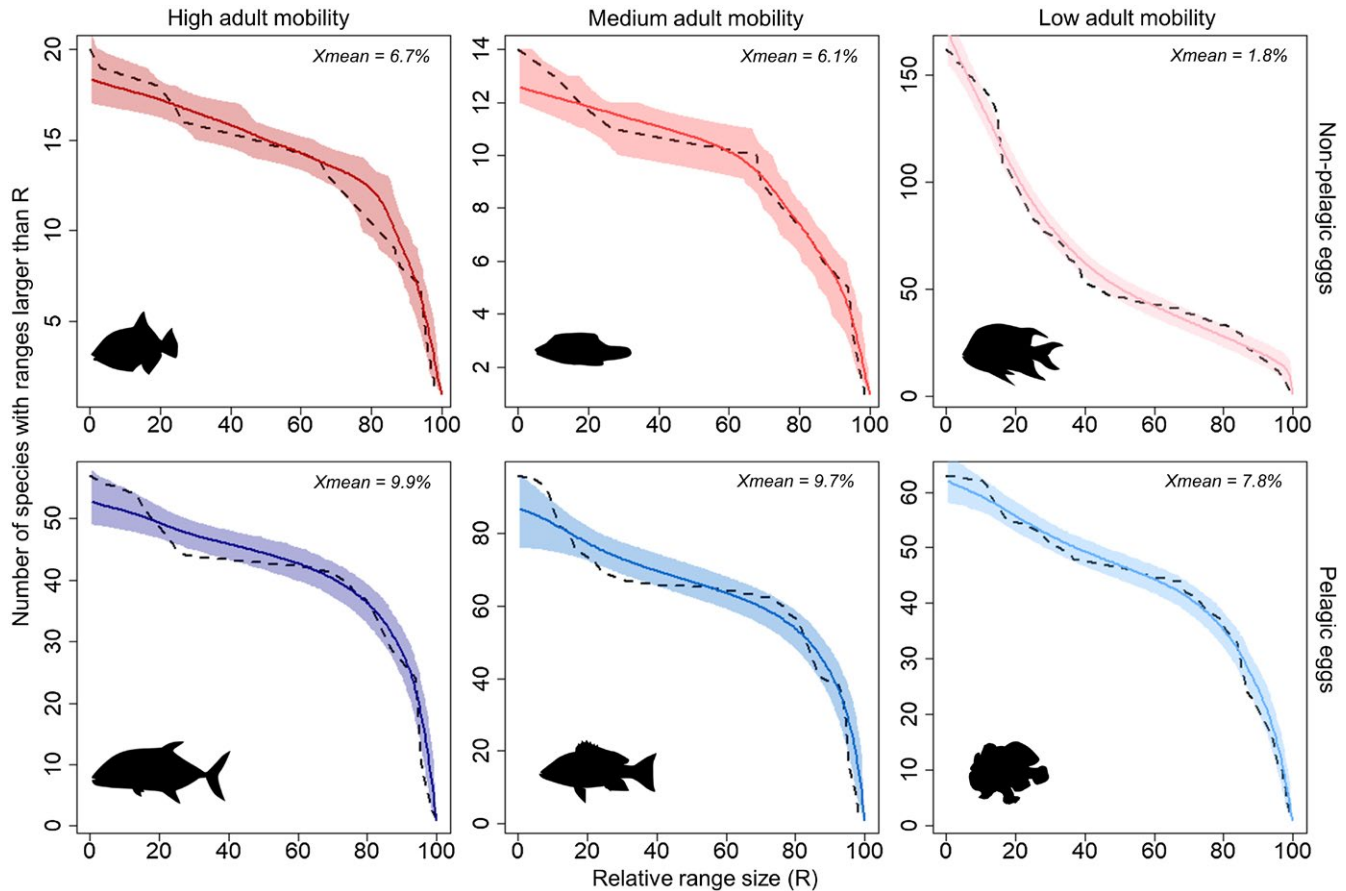


**FIGURE 5** Effect on range size distributions of (a) the mean dispersal distance  $X_{mean}$ , (b) the shape parameter of the dispersal kernel  $\alpha$ , (c) speciation  $\mu$ , (d) the time to speciation  $\tau$  (protractedness), (e) habitat fragmentation with  $X_{mean} = 40\%$  (top curves) and  $X_{mean} = 2\%$  (bottom curves) and (f) sampling proportion  $s$

Notes. Lines show the average value of 100 replicates and the shaded areas represent the 95% Confidence intervals (CI). For all simulations, we used a contiguous lattice with a size of 50,000 individuals. We use one fixed parameter setting, and varied only one variable of interest at a time:  $s = 100\%$ ,  $\alpha = 3.0$ ,  $X_{mean} = 0.02$ ,  $\mu = 0.0005$ ,  $\tau = 10$ .

models like the mid-domain effect model or the spreading dye model can provide null expectations of species richness patterns in absence of environmental gradients and under geometric constraints (Colwell & Lees, 2000; Jetz & Rahbek, 2001). However, these models are not able to provide predictions on range size distributions and have received criticism not only because of the use of observed ranges as a model input, but also because of the model assumptions (Zapata, Gaston, & Chown, 2003, 2005). Therefore, we created the

simplest non-mechanistic null model that can produce range size distributions. Our null model shows that under a scenario without dispersal limitation, species are expected to be more widely distributed and have larger ranges than they do in reality. This applies to all dispersal guilds, even those representing excellent overall dispersers. Our null model is thus unable to explain the range size distribution patterns in the TEP, suggesting that species distributions must also be affected by dispersal limitation. More sophisticated mechanistic



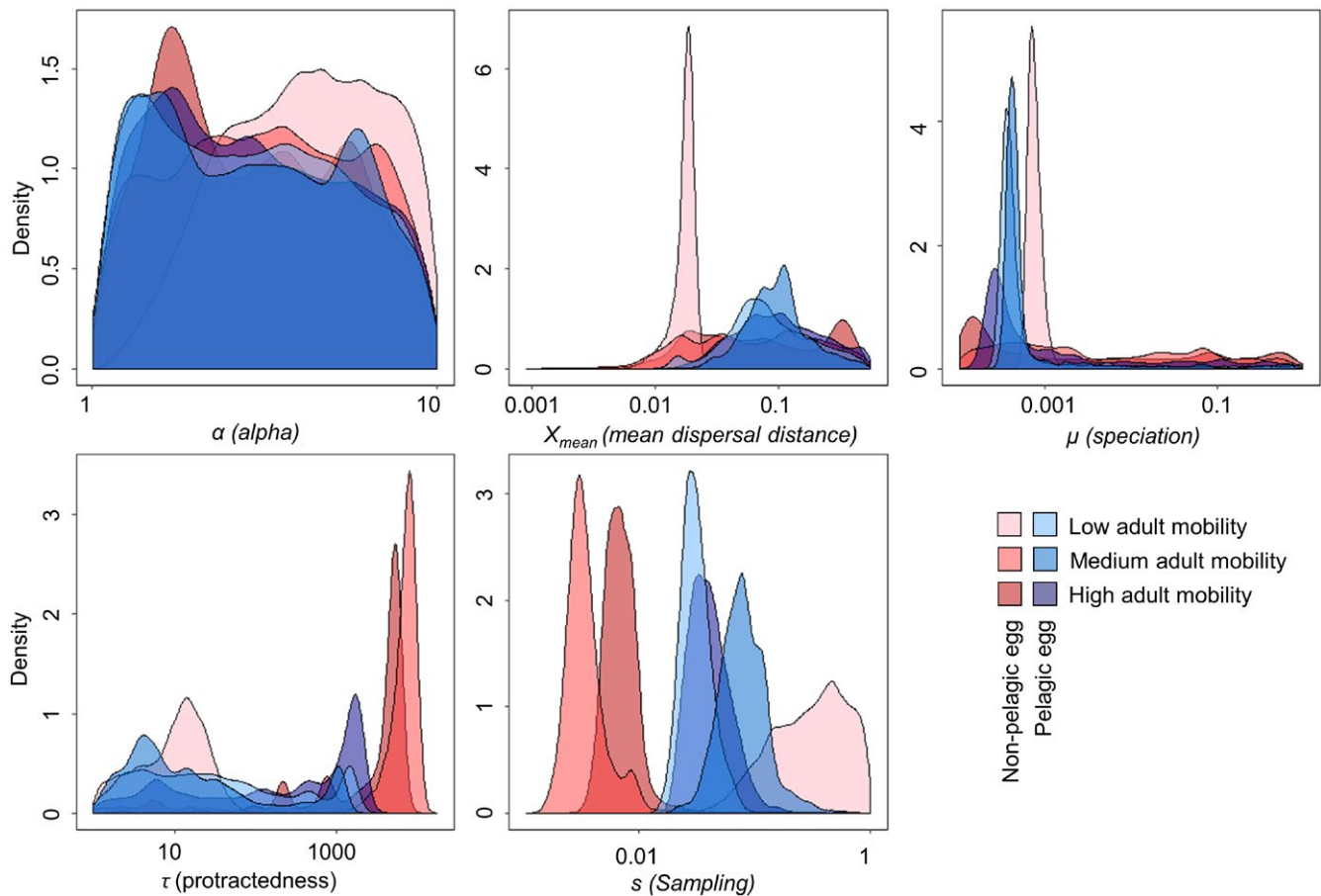
**FIGURE 6** Range size distributions of the best-fitting model for each dispersal guild, shown as an inverse cumulative distribution curve. Notes. Mean of five replicates and 95% Confidence intervals (CI) are shown. Dashed lines represent the empirical data and coloured bands represent the distribution of values in the best-fitting model for that guild. Estimated  $X_{\text{mean}}$  (median of > 90,000 estimates) is shown for each dispersal guild.

models (that include dispersal limitation) are therefore a next step to reproduce range size distribution patterns. The neutral model of biodiversity and biogeography (Hubbell, 2001) has proven to be a useful mechanistic model that can produce several biodiversity patterns, for example, species abundance distribution, species–area relationships and beta-diversity patterns, while at the same time remaining conceptually simple. Neutral models can be used as a tool to study biodiversity patterns (Rangel & Diniz-Filho, 2005; Tittensor & Worm, 2016). They have been shown to adequately reproduce fish diversity patterns in riverine ecosystems (Muneepeerakul et al., 2008), and longitudinal gradients in species richness, diversification rates and beta-diversity patterns in mangroves (Descombes et al., 2018).

Our spatially explicit model provides evidence that two traits used as a proxy of dispersal, spawning mode and adult mobility, are indeed related to dispersal ability and to range sizes in tropical reef fishes. The importance of dispersal ability in explaining range size variation has often been questioned, due to mixed results of several correlational studies (Lester & Ruttenberg, 2005; Luiz et al., 2013; Ruttenberg & Lester, 2015). Our model shows that high dispersal rates generate distributions where many species have large ranges, whereas low dispersal leads to the emergence of a large proportion of small ranged species, consistent with a positive relationship between

dispersal and range size. Interestingly, our model also shows that range size variation can be large within dispersal guilds, as dispersal only affects the *probability* of long-distance dispersal. Thus, although many species with low dispersal have small ranges, there are some with large ranges. Similarly, high dispersive species sometimes have small ranges despite large ranges being the norm. These findings help explain why it has been challenging for empirical and correlative studies to uncover clear links between dispersal ability and range size.

Our neutral model predicts range size distributions with a close fit to the empirical distributions for six different dispersal guilds of reef fishes in the TEP, and for each guild, estimated mean dispersal distance was in line with expectations for that guild's dispersal ability. These findings indicate that, despite their simplicity, neutral models still capture the most important processes for driving range size variation within such guilds (although of course not between guilds with different dispersal characteristics). Within guilds of pelagic spawners with high and medium adult mobility, the range size distribution tended to be bimodal, which cannot be explained by neutral processes alone. We found that this bimodality primarily resulted from the combination of two different background distributions: TEP endemics and TEP non-endemics, with the endemics generally having larger ranges within the TEP. We hypothesize that



**FIGURE 7** Posterior distribution of parameters for each guild, inferred using approximate Bayesian computation

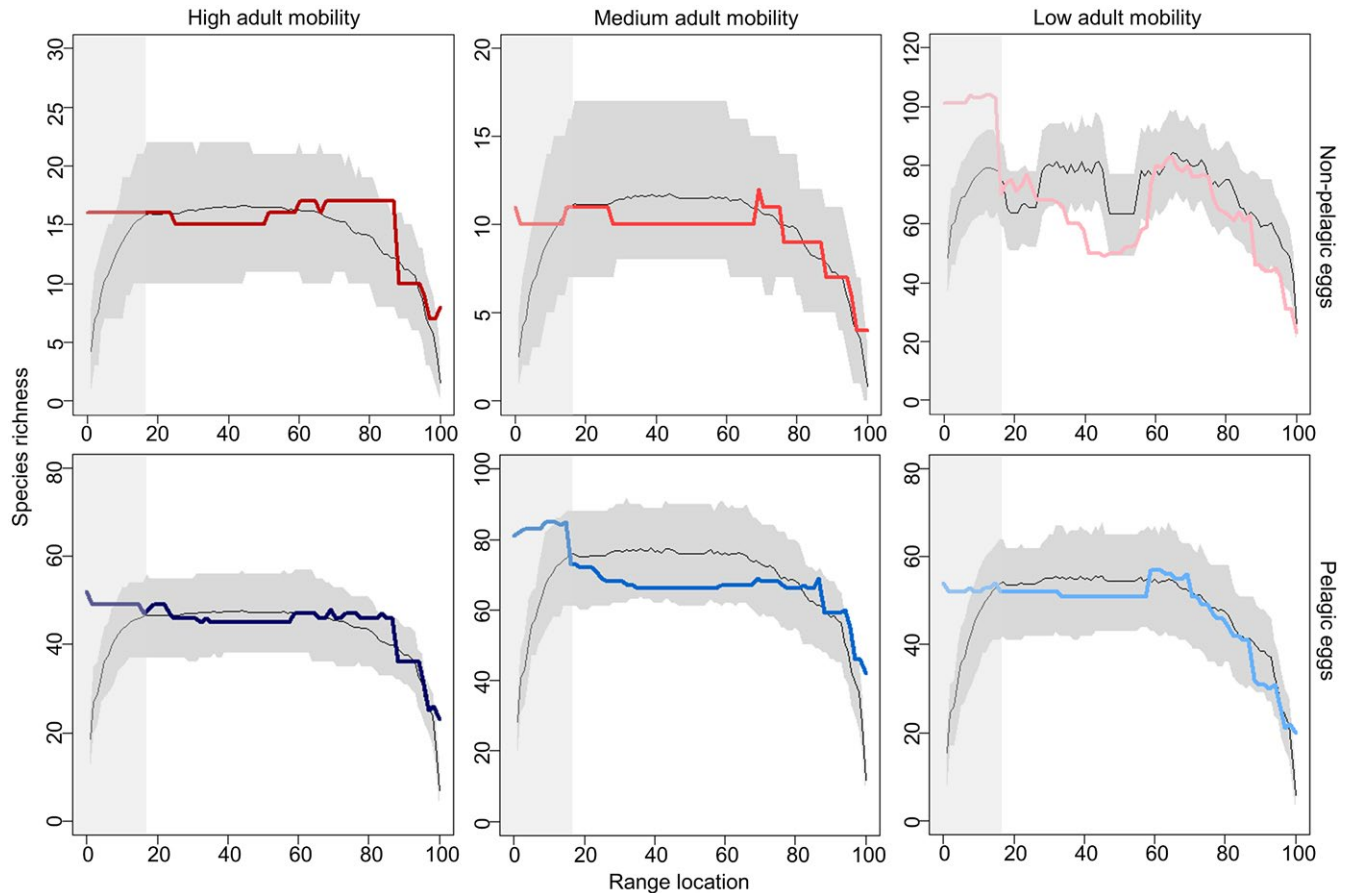
Notes. Shown are results for the shape of the dispersal kernel ( $\alpha$ ), mean dispersal distance ( $X_{\text{mean}}$ ), speciation initiation rate ( $\mu$ ), protractedness ( $\tau$ ) and the rate of sampling ( $s$ ). Whereas the pelagic guilds generally have very similar parameter distributions regardless of mobility, the non-pelagic guilds differ strongly, with the lowest adult mobility guild standing out.

endemics have generally had a longer time to increase their ranges in the region. In contrast, non-endemics include recent immigrant populations and have both large and small ranges; they are often just at the edges of the range for a wider ranged species that mostly occupies areas outside the TEP. Possibly, such non-endemic species migrated from temperate regions (North and South America), and others from tropical areas outside the TEP, so they are climatically constrained from expanding further in ways that cannot be captured by a neutral model. We conjecture that species origin has a major influence on range expansion via successful colonization, as a consequence of their adaptation to tropical conditions. For instance, the majority of non-endemic species with large ranges are trans-Pacific species, already adapted to tropical conditions. In contrast, 22 out of the 24 species with very small ranges come from temperate regions, and it is likely that their adaptations to a temperate climate made these species less able to expand their ranges into tropical regions (Holt, 2003). Species coming from the temperate north indeed do not go down to the south and vice versa, whereas transpacific species are well distributed along the coast (Supporting Information Appendix S10). This supports our conjecture of climatic constraints playing a role for the non-endemic species.

Speciation, sampling intensity and dispersal are major determinants of range size formation in our study. When sampling effort was low, only a single individual was detected for many species (hence they were treated as singletons, even though more individuals may have been present but not observed), leading to a high proportion of species with very small ranges. The proportion of species with small ranges also increased when speciation rates were high, or when speciation was a fast process (low protractedness). In these cases, new species emerged continuously with low abundance and restricted range. Empirical data also show that the lowest dispersive guild (low adult mobility and non-pelagic spawners) has a lot more species than more dispersive guilds. The interaction between low dispersal and geographical isolation may facilitate speciation via reduction in gene flow (Riginos, Buckley, Blomberg, & Tremblay, 2014), and hence lead to higher diversity. In our model, speciation has the same effect without the inclusion of reduction of gene flow influencing speciation: low dispersal guilds tend to have more species than high dispersive guilds, and habitat fragmentation can further strengthen this pattern (Supporting Information Appendix S11).

In addition to the most obvious processes related to range size, our model also shows that habitat fragmentation can play





**FIGURE 8** Richness gradients predicted by the neutral model

**Notes.** Black solid lines and grey bands show the mean and the 95% CI of 100 simulations. Simulations used as an input the estimated parameter values estimated by fitting the model with fragmentation to the range size distribution of reef fishes (Supporting Information Appendix S8). Coloured curves show the empirical richness gradient along the TEP coast, the shaded areas highlight the range locations affected by collapsing the Gulf of California (see methods), which unrealistically increases local species richness on those localities. Locations along the coastline are transformed to relative number.

an important role in shaping macroecological patterns and that its effect depends on the dispersal abilities of the species. For instance, highly dispersive species have range size distributions that do not differ much between contiguous and fragmented habitats. Conversely, range size distributions for low dispersive species are strongly affected by habitat fragmentation. These species tend to have even smaller ranges in fragmented habitats than in contiguous habitats. This suggests that our model's predictions conform to the intuition that range expansion for low dispersive species is negatively affected by habitat fragmentation.

By running simulations with parameters estimated from range size distributions (richness patterns were not fitted), our neutral model can predict species richness gradients in the region reasonably well. Interestingly, the predicted richness gradients depend on dispersal and its interaction with habitat fragmentation. The predicted species richness gradients along the TEP coast tend to have the typical convex shape (richness increasing towards the centre of the distribution gradient) for highly dispersive guilds, whereas for the lowest dispersive guild (low adult mobility and non-pelagic spawners), the shape was quite different from the other guilds and

looks like a more complex form with several interior peaks (Figure 8). In line with Macpherson, Hastings, and Robertson (2009), broadly distributed species (which tend to be the more dispersive) in the TEP are responsible for the convex shape of the latitudinal richness gradient, whereas narrow-ranged species (which tend to be species with low adult mobility and non-pelagic spawners) are responsible for departures from the typical convex shape. Narrowly distributed species are concentrated in two stretches of coast where there is more contiguous reef habitat (Gulf of California and Panama/Costa Rica). Exploration of our model showed that the predicted richness gradients for low dispersive species in a contiguous habitat will be a convex curve and that departures from that curve (bimodality or other shapes) are caused by habitat fragmentation and inability of those species to cross the dispersal barriers (Supporting Information Appendix S11).

Here, we have shown that dispersal is a crucial factor in shaping the range size distribution of species. The use of both a mechanistic model and detailed analyses was required to reach this conclusion. Variation in range size across species can be explained by a combination of neutral processes and guild-specific differences



in dispersal. Consistent with previous studies on neutral models with guild structure (using predictions for abundance instead of range size, Aduse-Poku et al., 2018; Janzen, Haegeman, & Etienne, 2015), our results show that while community dynamics within guilds may be captured by a neutral model, across guilds niche-based processes drive variation in range size. Neutral theory was originally proposed to describe community assembly within guilds (Hubbell, 2001). Our results are consistent with this philosophy but take the concept further by fitting both a single neutral model and a set of independent neutral guilds to empirical data (Supporting Information Appendix S12, Figure 6). Our results show that across guilds, niche-based processes, in this case differing dispersal strategies, play a larger role in driving ecological patterns, while within guilds they are much less important. The neutral models we used were originally developed to understand macroecological patterns such as species abundances and species–area relationships (Hubbell, 2001). Neutral models can thus be regarded as a generic mechanistic tool, which we apply here to new patterns, rather than a phenomenological construct tailored to range sizes alone. Our findings thereby make substantial progress towards settling the long-standing question of what causes variation in range size, and of the role of dispersal in this pattern.

## ACKNOWLEDGMENTS

We thank two anonymous referees for helpful comments improving the manuscript. We thank Fons van der Plas for comments on previous versions of the manuscript. We thank Dr Ross Robertson for providing us with information about the distribution of reefs in the Tropical Eastern Pacific. A.A. was funded by the Ubbo Emmius Fund and by the BelSpo IAP project “SPatial and environmental determinants of Eco-Evolutionary DYnamics: anthropogenic environments as a model”. T.J. is grateful for use of the computational cluster ADA of the Max Planck Institute for Evolutionary Biology, Plön, and for the use of the computational cluster Peregrine, of the University of Groningen. J.R. was funded by fellowships from the Natural Environment Research Council (NERC) (NE/I021179, NE/L011611/1). Through J.R., this study is a contribution to Imperial College's Grand Challenges in Ecosystems and the Environment initiative. D.B. and R.E. received support from the FWO research community an eco-evolutionary network of biotic interactions (EVENET) and FWO project G.018017.N. R.S.E. received support by a Innovational Research Incentives Scheme Vici (VICI) grant (865.13.003) from the Netherlands Organisation for Scientific Research (NWO).

## DATA ACCESSIBILITY

Data on range size of reef fishes used in this study and the spatially explicit neutral model are available on the DataVerseNL digital repository <https://hdl.handle.net/10411/VMQPCW>. The code for the spatially explicit neutral model is available on GitHub [https://github.com/thijsjanzen/Neutral\\_Range\\_Size/](https://github.com/thijsjanzen/Neutral_Range_Size/).

## ORCID

Adriana Alzate  <https://orcid.org/0000-0002-8236-6510>

Thijs Janzen  <https://orcid.org/0000-0002-4162-1140>

Dries Bonte  <https://orcid.org/0000-0002-3320-7505>

Rampal S. Etienne  <https://orcid.org/0000-0003-2142-7612>

## REFERENCES

- Addis, D. T., Patterson, W. F. III, Dance, M. A., & Ingram, G. W. Jr. (2013). Implication of reef fish movement from unreported artificial reef sites in the northern Gulf of Mexico. *Fisheries Research*, 147, 349–358. <https://doi.org/10.1016/j.fishres.2013.07.011>
- Aduse-Poku, K., Molleman, F., Oduro, W., Oppong, S. K., Lohman, D. J., & Etienne, R. S. (2018). Relative contribution of neutral and deterministic processes in shaping fruit-feeding butterfly assemblages in Afrotropical forests. *Ecology and Evolution*, 8, 296–308. <https://doi.org/10.1002/ece3.3618>
- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6), 716–723. <https://doi.org/10.1109/TAC.1974.1100705>
- Alzate, A., van der Plas, A., Zapata, F. A., Bonte, D., & Etienne, R. S. (2019). Incomplete datasets obscure associations between traits affecting dispersal ability and geographic range size of reef fishes in the Tropical Eastern Pacific. *Ecology and Evolution*. <https://doi.org/10.1002/ece3.4734>
- Alzate, A., Zapata, F. A., & Giraldo, A. (2014). A comparison of visual and collection-based methods for assessing for community structure of coral reef fishes in the Tropical Eastern Pacific. *Revista de Biología Tropical*, 62(1), 359–371. <https://doi.org/10.15517/rbt.v62i0.16361>
- Anderson, S. (1985). The theory of range-size (RS) distributions. *American Museum Novitates*, 2833, 1–20. <http://hdl.handle.net/2246/3587>
- Appeldoorn, R. S., Hensley, D. A., Shapiro, D. G., Kioroglou, S., & Sanderson, B. G. (1994). Egg dispersal in a Caribbean coral reef fish, *Thalassoma bifasciatum*. II. Dispersal off the reef platform. *Bulletin of Marine Science*, 54(1), 271–280.
- Beaumont, M. A., Zhang, W., & Balding, D. J. (2002). Approximate Bayesian computation in population genetics. *Genetics*, 162(4), 2025–2035.
- Bonhomme, F., & Planes, S. (2000). Some evolutionary arguments about what maintains the pelagic interval in reef fishes. *Environmental Biology of Fishes*, 59(4), 365–383.
- Bonte, D., Van Dyck, H., Bullock, J. M., Coulon, A., Delgado, M., Gibbs, M., ... Travis, J. M. J. (2012). Costs of dispersal. *Biological Reviews*, 87, 290–312. <https://doi.org/10.1111/j.1469-185X.2011.00201.x>
- Brown, J. H., & Kondric-Brown, A. (1977). Turnover rates in insular biogeography: Effects of immigration on extinction. *Ecology*, 58, 445–449. <https://doi.org/10.2307/1935620>
- Burgess, S. C., Baskett, M. L., Grosberg, R. K., Morgan, S. G., & Strathmann, R. R. (2015). When is dispersal for dispersal? Unifying marine and terrestrial perspectives. *Biological Reviews*, 91(3), 867–882. <https://doi.org/10.1111/brv.12198>
- Colwell, R. K., & Lees, D. C. (2000). The mid-domain effect: Geometric constraints on the geography of species richness. *Trends in Ecology and Evolution*, 15, 70–76. [https://doi.org/10.1016/S0169-5347\(99\)01767-X](https://doi.org/10.1016/S0169-5347(99)01767-X)
- Connolly, S. R., Keith, S. A., Colwell, R. K., & Rahbek, C. (2017). Processes, mechanisms, and modeling in macroecology. *Trends in Ecology and Evolution*, 32, 835–844. <https://doi.org/10.1016/j.tree.2017.08.011>
- Csilléry, K., François, O., & Blum, M. G. (2012). abc: An R package for approximate Bayesian computation (ABC). *Methods in Ecology and Evolution*, 3, 475–479. <https://doi.org/10.1111/j.2041-210X.2011.00179.x>

- Dennis, R. L. H., Sparks, T. H., & Hardy, B. P. (1999). Bias in butterfly distribution maps: The effect of sampling effort. *Journal of Insect Conservation*, 3, 33–42. <https://doi.org/10.1023/A:1009678422145>
- Descombes, P., Gaboriau, T., Albouy, C., Heine, C., Leprieux, F., & Pellissier, L. (2018). Linking species diversification to paleo-environmental changes: A process-based modelling approach. *Global Ecology and Biogeography*, 27(2), 233–244. <https://doi.org/10.1111/geb.12683>
- Etienne, R. S., & Rosindell, J. (2012). Prolonging the past counteracts the pull of the present: Protracted speciation can explain observed slow-downs in diversification. *Systematic Biology*, 61, 204–2013. <https://doi.org/10.1093/sysbio/syr091>
- Floeter, S. R., Ferreira, C. E. L., Dominici-Arosemena, A., & Zalmon, I. R. (2004). Latitudinal gradients in Atlantic reef fish communities: Trophic structure and spatial use patterns. *Journal of Fish Biology*, 64, 1680–1699. <https://doi.org/10.1111/j.0022-1112.2004.00428.x>
- Gaston, K. J. (1996). Species-range-size distributions: Patterns, mechanisms and implications. *Trends in Ecology and Evolution*, 11, 197–201. [https://doi.org/10.1016/0169-5347\(96\)10027-6](https://doi.org/10.1016/0169-5347(96)10027-6)
- Gaston, K. J. (2003). *The structure and dynamics of geographic ranges*. New York, NY: Oxford University Press.
- Gaston, K. J., & Chown, S. L. (1999). Geographic range size and speciation. In A. E. Magurran & R. M. May (Eds.), *Evolution of biological diversity* (pp. 236–259). UK: Oxford University Press.
- Gaston, K. J., & He, F. (2002). The distribution of species range size: A stochastic process. *Proceedings of the Royal Society B: Biological Sciences*, 269, 1079–1086. <https://doi.org/10.1098/rspb.2002.1969>
- Hanski, I. (1982). Dynamics of regional distribution: The core and satellite species hypothesis. *Oikos*, 38, 210–221. <https://doi.org/10.2307/3544021>
- Harfoot, M. B. J., Newbold, T., Tittensor, D. P., Emmott, S., Hutton, J., Lyutsarev, V., ... Purves, D. W. (2014). Emergent global patterns of ecosystem structure and function from a mechanistic general ecosystem model. *PLoS Biology*, 12(4), e1001841. <https://doi.org/10.1371/journal.pbio.1001841>
- Holt, R. D. (2003). On the evolutionary ecology of species' ranges. *Evolutionary Ecology Research*, 5, 159–178.
- Holt, R. D., & Gomulkiewicz, R. (1996). The evolution of species' niches: a population dynamic perspective. In H. G. Othmer, F. R. Adler, M. A. Lewis, & L. C. Dallon (Eds.), *Case studies in mathematical modelling* (pp. 25–50). Saddle River, NJ: Prentice-Hall.
- Hubbell, S. P. (2001). *The unified neutral theory of biodiversity and biogeography*. NJ: Princeton University Press.
- Janzen, T., Haegeman, B., & Etienne, R. S. (2015). A sampling formula for ecological communities with multiple dispersal syndromes. *Journal of Theoretical Biology*, 374, 94–106. <https://doi.org/10.1016/j.jtbi.2015.03.018>
- Jetz, W., & Rahbek, C. (2001). Geometric constraints explain much of the species richness pattern in African birds. *Proceedings of the National Academy of Sciences USA*, 98, 5661–5666. <https://doi.org/10.1073/pnas.091100998>
- Jones, G. P. (2015). Mission impossible: unlocking the secrets of coral reef fish dispersal. In C. Mora (Ed.), *Ecology of fishes on coral reefs* (pp. 16–27). UK: Cambridge University Press.
- Kaunda-Arara, B., & Rose, G. A. (2004). Long-distance movements of coral reef fishes. *Coral Reefs*, 23, 410–412. <https://doi.org/10.1007/s00338-004-0409-7>
- Leis, J. M. (1978). Systematics and zoogeography of the porcupinefishes (Diodon, Diodontidae, Tetraodontiformes), with comments on egg and larvae development. *Fishery Bulletin*, 76(3), 535–567.
- Leis, J. M. (2006). Are larvae of demersal fishes plankton or nekton. *Advances in Marine Biology*, 51, 57–141. [https://doi.org/10.1016/S0065-2881\(06\)51002-8](https://doi.org/10.1016/S0065-2881(06)51002-8)
- Leis, J. M., Caselle, J. E., Bradbury, I. R., Kristiansen, T., Llopiz, J. K., Miller, M. J., ... Warner, R. R. (2013). Does fish larvae dispersal differ between high and low latitudes? *Proceedings of the Royal Society B: Biological Sciences*, 280, 20130327. <https://doi.org/10.1098/rspb.2013.0327>
- Lester, S. E., & Ruttenberg, B. I. (2005). The relationship between pelagic larval duration and range size in tropical reef fishes: A synthetic analysis. *Proceedings of the Royal Society B: Biological Sciences*, 72, 585–591. <https://doi.org/10.1098/rspb.2004.2985>
- Lester, S. E., Ruttenberg, B. I., Gaines, S. D., & Kinlan, B. P. (2007). The relationship between dispersal ability and geographic range size. *Ecology Letters*, 10, 745–758. <https://doi.org/10.1111/j.1461-0248.2007.01070.x>
- Luiz, O. J., Allen, A. P., Robertson, D. R., Floeter, S. R., Kulbicki, M., Vigliola, L., ... Madin, J. S. (2013). Adult and larval traits as determinants of geographic range size among tropical reef fishes. *Proceedings of the National Academy of Sciences, USA*, 110(41), 16498–16502. <https://doi.org/10.1073/pnas.1304074110>
- MacArthur, R. H., & Wilson, E. O. (1967). *The theory of island biogeography*. NJ: Princeton University Press.
- Macpherson, E., Hastings, P. A., & Robertson, D. R. (2009). Macroecological patterns among marine fishes. In J. D. Witman & K. Roy (Eds.), *Marine macroecology* (pp. 122–152). IL: The University of Chicago Press.
- Magnusson, A. (2015). *gmt: Interface between GMT map-making software and R*. R package version 1.2-0. Retrieved from <http://CRAN.R-project.org/package=gmt>
- Mora, C., Trembl, E., Robert, J., Crosby, K., Roy, D., & Tittensor, D. P. (2012). High connectivity among habitats precludes the relationship between dispersal and range size in tropical reef fishes. *Ecography*, 35(1), 89–96. <https://doi.org/10.1111/j.1600-0587.2011.06874.x>
- Munday, P. L., & Jones, G. P. (1998). The ecological implications of small body size among coral-reef fishes. *Oceanography and Marine Biology: An Annual Review*, 36, 373–411.
- Muneepeerakul, R., Bertuzzo, E., Lynch, H. J., Fagan, W. F., Rinaldo, A., & Rodriguez-Iturbe, I. (2008). Neutral metacommunity models predict fish diversity patterns in Mississippi-Missouri basin. *Nature*, 453, 220–223. <https://doi.org/10.1038/nature06813>
- R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org>
- Rangel, T. F. L. V. B., & Diniz-Filho, J. A. F. (2005). Neutral community dynamics, the mid-domain effect and spatial patterns in species richness. *Ecology Letters*, 8, 783–790. <https://doi.org/10.1111/j.1461-0248.2005.00786.x>
- Riginos, C., Buckley, Y. M., Blomberg, S. P., & Trembl, E. A. (2014). Dispersal capacity predicts both population genetic structure and species richness in reef fishes. *The American Naturalist*, 184(1), 54–64. <https://doi.org/10.1086/676505>
- Riginos, C., Douglas, C. E., Jin, Y., Shanahan, D. F., & Trembl, E. A. (2011). Effects of geography and life history traits on genetic differentiation in benthic marine fishes. *Ecography*, 34, 566–575. <https://doi.org/10.1111/j.1600-0587.2010.06511.x>
- Robertson, D. R., & Allen, G. (2016). *Fishes: East Pacific. An Identification Guide to the Shore-Fish Fauna of the Tropical Eastern Pacific*. (Copyright Smithsonian Institution, Left Coast R&C, Santa Cruz, California). Retrieved from <http://biogeodb.stri.si.edu/bioinformatics/sfstep/taxon.php>
- Rosindell, J., Cornell, S. J., Hubbell, S. P., & Etienne, R. S. (2010). Protracted speciation revitalizes the neutral theory of biodiversity. *Ecology Letters*, 13(6), 716–727. <https://doi.org/10.1111/j.1461-0248.2010.01463.x>
- Rosindell, J., Wong, Y., & Etienne, R. S. (2008). A coalescence approach to spatial neutral ecology. *Ecological Informatics*, 3(3), 259–271. <https://doi.org/10.1016/j.ecoinf.2008.05.001>
- Ruttenberg, B. I., & Lester, S. E. (2015). Patterns and processes in geographic range size in coral reef fishes. In C. Mora (Ed.), *Ecology of fishes on coral reefs* (pp. 97–103). UK: Cambridge University Press.

- Stahl, U., Reu, B., & Wirth, C. (2014). Predicting species range limits from functional traits for the tree flora of North America. *Proceedings of the National Academy of Sciences USA*, 111, 13739–13744. <https://doi.org/10.1073/pnas.1300673111>
- Stobutzki, I. C., & Bellwood, D. R. (1997). Sustained swimming abilities of the late pelagic stages of coral reef fishes. *Marine Ecology Progress Series*, 149, 35–41. <https://doi.org/10.3354/meps149035>
- Tittensor, D. P., & Worm, B. (2016). A neutral-metabolic theory of latitudinal biodiversity. *Global Ecology and Biogeography*, 25(6), 630–641. <https://doi.org/10.1111/geb.12451>
- Toni, T., Welch, D., Strelkowa, N., Ipsen, A., & Stumpf, M. P. H. (2009). Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *Journal of the Royal Society Interface*, 6, 187–202. <https://doi.org/10.1098/rsif.2008.0172>
- Webb, T. J., & Gaston, K. J. (2000). Geographic range size and evolutionary age in birds. *Proceedings of the Royal Society B: Biological Sciences*, 267(1455), 1843–1850. <https://doi.org/10.1098/rspb.2000.1219>
- Wootton, R. J. (1992). *Fish ecology*. New York, NY: Chapman & Hall.
- Zapata, F. A., Gaston, K. J., & Chown, S. L. (2003). Mid-domain models of species richness gradients: Assumptions, methods and evidence. *Journal of Animal Ecology*, 72, 677–690. <https://doi.org/10.1046/j.1365-2656.2003.00741.x>
- Zapata, F. A., Gaston, K. J., & Chown, S. L. (2005). The mid-domain effect revisited. *The American Naturalist*, 166(5), e144–e148. <https://doi.org/10.1086/491685>

## BIOSKETCH

ADRIANA ALZATE is interested in ecology, evolution and macroecology, particularly in how eco-evolutionary dynamics can shape macroecological patterns.

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

**How to cite this article:** Alzate A, Janzen T, Bonte D, Rosindell J, Etienne RS. A simple spatially explicit neutral model explains the range size distribution of reef fishes. *Global Ecol Biogeogr*. 2019;00:1–16. <https://doi.org/10.1111/geb.12899>